



DOMA – relevant existing activities

Oliver Keeble

Overview

- Current WLCG initiatives
- Historical WLCG initiatives
- Current related activities
- Operations

Current

- Archival storage
- Storage resource reporting
- DM steering

Archival Storage WG

- Two ongoing initiatives
 - Establishing best practise for use of tape systems
 - in preparation for more intensive use of archival systems
 - Identifying and reporting metrics
- See Vlado's HEPiX talk
 - <https://indico.cern.ch/event/676324/contributions/2967985/>
- Status
 - Site survey complete, formulating recommendations
 - First metrics defined, encouraging publication
- Future
 - Exercising new workflows, e.g. carousels
 - Incorporation of QoS
 - Generalisation of “tape”

Resource reporting

- Produced a set of recommendations
 - To support WLCG storage accounting and experiment operations
 - How to get used and free space
 - without relying on SRM.
 - https://twiki.cern.ch/twiki/bin/view/LCG/AccountingTaskForce#Storage_Space_Accounting
 - Connection to InfoSys evolution and CRIC
- Status
 - Recommendations implemented by a number of storage systems
 - Rollout far from complete
 - Work on accounting portal progressing
- Future
 - Full exposure of storage costs
 - QoS

Data Management Steering

- Medium/Long term strategy
 - <https://twiki.cern.ch/twiki/bin/view/LCG/WLCGDataSteeringGroup>
- In practise – decide into which topics common effort should be invested
- Produced a report on priorities and organised a number of meetings
 - preGDBs, WLCG workshop sessions
- Future
 - Should be absorbed into the DOMA activity at the strategic level.

Historical interest

- HTTP Deployment TF
 - Deployment advice and SAM probes
- Caching “interest group”
- preGDBs
 - <https://indico.cern.ch/category/6890/>
 - Object storage
 - Protocol Zoo
 - Archival storage
 - Resource reporting
- TEGs
 - <https://indico.cern.ch/category/3772/>
 - Storage Management
 - Data Management
 - Access patterns
 - Protocol zoo
- Federated Data Stores

Related

- HTTP Transfer Ecosystem
 - Demonstration of HTTP data distribution
 - Scitokens project
 - Delegation free HTTP transfers
- XDC (<http://www.extreme-datacloud.eu/>)
 - Investing in dCache, Dynafed, EOS, FTS and Orchestration technologies to create data infrastructure software for WLCG and other communities
- Grid Community Forum (<http://gridcf.org>)
 - Now the guardians of GSI and GridFTP
- HEPiX WGs
 - <https://indico.cern.ch/category/5233/>
- HEP Software Foundation
 - In particular the Community White Paper on DOMA
 - <https://hepsoftwarefoundation.org/activities/cwp.html>

Beyond DOMA

- “DOMA” encourages a global view of data related activities
 - But one can zoom out further – we should connect (at least) to the following:
- Network
 - Data infrastructures are enabled by their networks
 - Crucial to develop strategies consistent with projected network evolution
 - <https://wlcg-ops.web.cern.ch/ops-coordination-working-groups/network-and-transfer-metrics>
 - HEPiX Network Functions Virtualisation WG
- AAI
 - Links many DOMA elements together
 - <https://twiki.cern.ch/twiki/bin/view/LCG/WLCGAuthorizationWG>
- Applications and I/O patterns
- Performance and Cost Modelling
- Data Preservation
- Workload Management...

Ops-ready topics

- Transfer protocol comparison
 - Identify implementation gaps and optimisation opportunities
- WebDAV deployment
 - Complete the process, make SAM probes critical
- SRM-free DM
 - Demonstrate that SRM is really optional
- Storage Accounting
 - Deployment of resource reporting recommendations
 - In particular, storage summary json file
- Regional storage consolidation
 - This is happening, is there enough knowledge sharing?
- Analysis of cached access
 - Compare strategies and analyse workflows